Sentiment Analysis of User Generated Online Content to Detect Suicidal Tendencies

-YUVRAJ SINGH

2013589

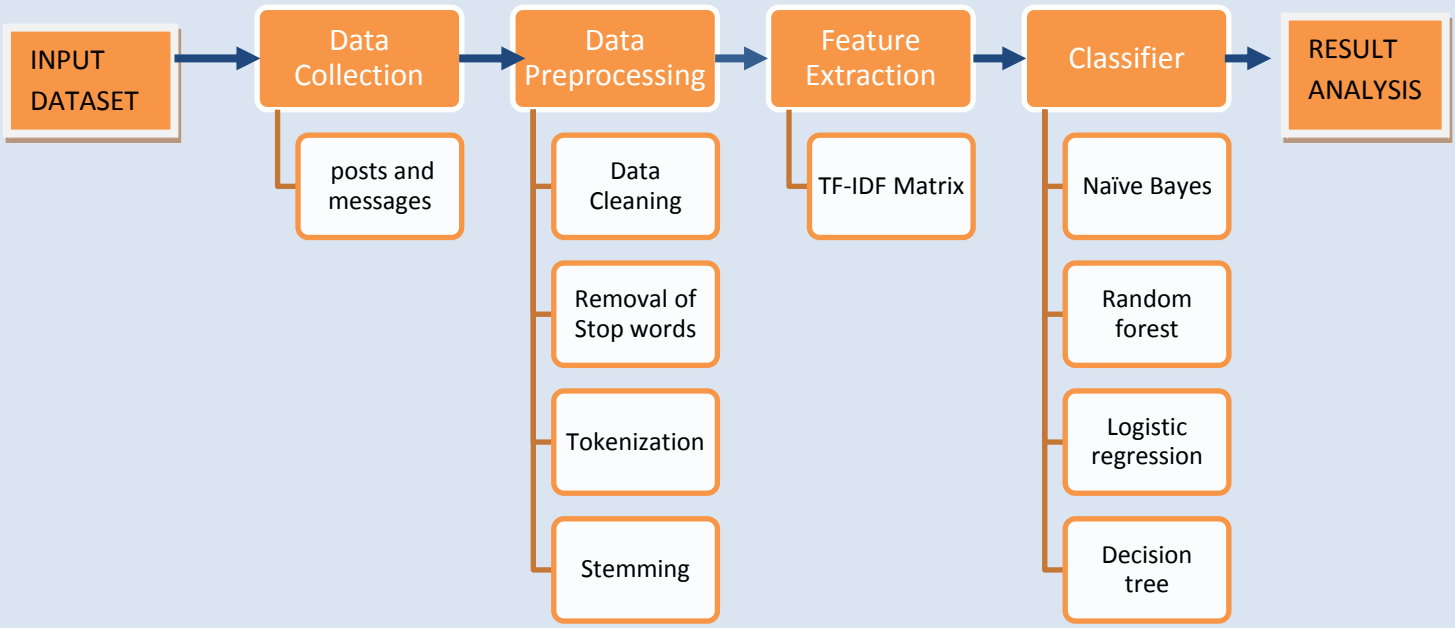
PROBLEM STATEMENT

- To detect suicidal ideation by processing the post uploaded by users on the Internet.
- Comparison of supervised learning algorithms for text classification for better accuracy

Motivation

The sudden demise of promising Bollywood actor Sushant Singh Rajput was a shock beyond words for me. He was a powerful performer that always motivated me to achieve great things in my life. According to reports he was suffering from depression from last few months and very few people knew about it, even in his afterlife he is still a source of motivation for me to help people suffering from this disorder.

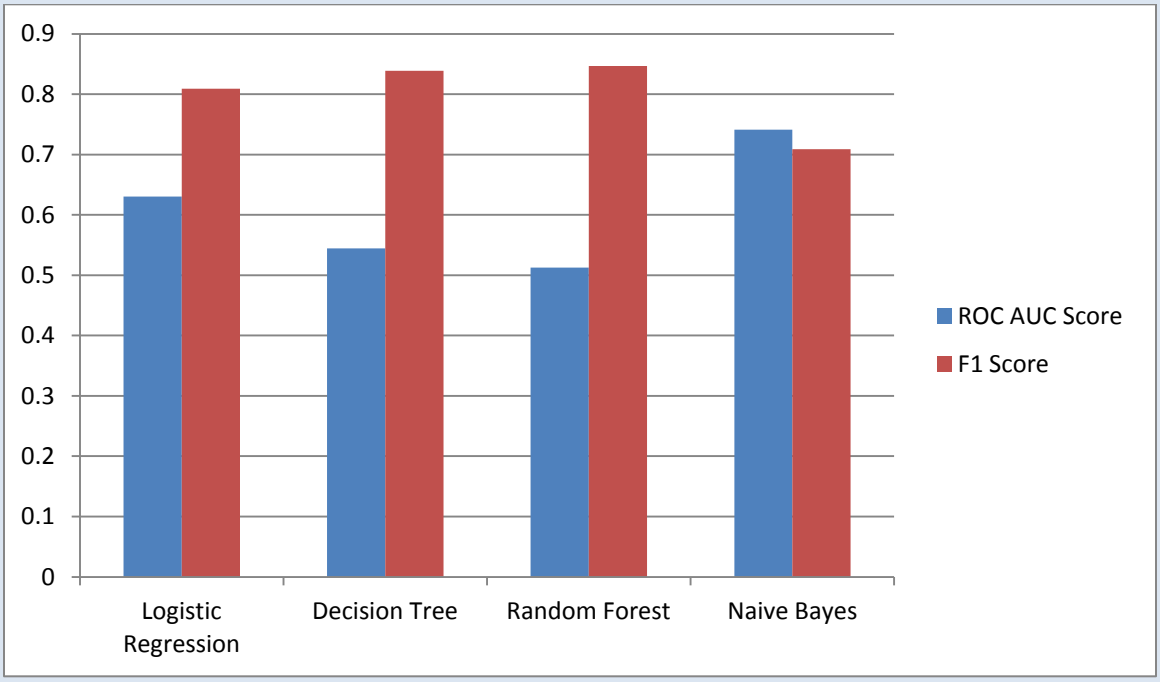
Methodology



- **In this project I used multiple classifier to check which of them gives me the best result, for the purpose of feature extraction I used (TF-IDF Matrix) Term Frequency-Inverse Document Frequency**
- **TF-IDF penalizes the words like 'don't', 'can't', and 'use' because they are commonly occurring words.**
- **For Data cleaning I used some basics data preprocessing Methods**

1. **Tokenization**
2. **Stemming**
3. **Lemmatization**

Performance Comparison



CONCLUSION

- **Random forest gave the best performance for TF-IDF feature set. It reaffirms the importance of probabilistic view for text classification.**
- **Accuracies for TF-IDF term frequency and Inverse document frequency is a better measure of similarities in texts.**
- **As per the results, count of False-Negatives is less than False-Positives for most of the algorithms indicating lesser count of suicidal tweets classified as non-suicidal.**